A METHOD AND ARRANGEMENT FOR AN IMPROVED BUFFER SOLUTION WITHIN A COMMUNICATION NETWORK SWITCH

Field of the Invention

The present invention relates to a switch and a method for organizing dataflow through a switch within packet switched or bursts switched networks. More particularly it relates to a switch within a telecommunication or data communication network, where the switch includes one or more outputs and a buffer unit, further it relates to a method for organizing dataflows in a telecommunication or data communication network including at least one switch where said switch is associated with at least one buffer.

Background of the Invention

Despite the economic recession, network traffic continuous to grow. Optical Packet Switching (OPS) is a promising candidate for a future cost effective network, supporting both high throughput and utilization. Two main issues of interest in OPS are however optical synchronization and buffering. Recently, a number of works has focused on using asynchronous packet switching, thereby avoiding the optical synchronization unit [1], [2]. Because of the immaturity of optical memory, Fiber Delay Line (FDL) based optical buffering in combination with using the wavelength dimension for contention resolution (i.e. a packet can be forwarded on an arbitrarily wavelength leading to the destination), has been investigated [3], [2]. An alternative to FDL's is to use electronic memory with a limited number of buffer inputs [1,3]. In either case, buffer interfaces, consisting of FDL's or OE-converters (opto electronic converters), will represent a major cost factor for an optical packet switched system and should therefore be minimized [4].

Delay through an OPS network is negligible compared to transmission delay [1]. However, if the buffer has fewer input interfaces than the total number of switch inputs, only a fraction of the packets passing through the switch will be buffered and hence, delayed [1,4].

In an asynchronous system, if packets can be scheduled from the buffer to the outputs, without causing contention with new packets from the inputs, PLR (packet loss ratio) will be brought to a minimum limit. This approach is called Input Priority (IP), since scheduling priority is given to the new packets arriving at the input, instead of to the buffered packets (i.e. Buffer Priority (BP)). In a Slotted IP system, this is simple. At the start of a time slot, packets in the buffer are scheduled only if there are vacant wavelengths after scheduling the packets arriving at the input. In asynchronous VLP (Variable Length Packets) operation, packet arrival and duration is not predictable. A new packet can arrive at the input at any random moment after a packet was scheduled from the buffer, hence making total Asynchronous IP (AIP) impossible when the number of buffer ports is limited.

Further, in asynchronous optical Metro packet rings, the same problem as above arises. When aggregating new packets onto the ring in the access nodes, packets already at the ring may contend with new packets. A detection circuit combined with a delay, e.g. a Fibre Delay Line (FDL), may be applied to first detect and then delay packets before passing the access node. However, this calls for extra components, both detectors and FDLs, making the principle potentially expensive. Additionally, in order to avoid packet collision when new packets enters the Metro ring, the FDLs will need to delay the packets longer than the duration of the longest of the new packets that enters the ring. This will impose additional delay of the packets already in the packet ring.

In this patent application, with reference to simulations, the ability of the fixed and incremental FDL buffering schemes described in [4], as well as electronic buffering, to support applications with a high demand to packet sequence and PLR is described.

Thus it is obvious that an asynchronous OPS system with a good PLR is needed. According to the present invention these and other problems will be solved using an Asynchronous Input Priority algorithm (AIP3Q) for VLP, that sorts buffered packets into a number of Queues in accordance with the packet length distribution, according to the present invention The algorithm is optimized for systems using the wavelength dimension for contention resolution with the goal of optimizing the buffer performance with respect to minimizing Packet Loss Ratio (PLR) and packet reordering.

Summary of the Invention

It is an object of the present invention to provide an arrangement and a method avoiding the above described problems.

The features defined in the independent claims enclosed characterize this method.

In particular, the present invention provides a switch and/or a Metro access node within an asynchronous communication network, where the switch includes one or more outputs and a buffer unit, said buffer unit is either an integral part of the switch or an external part of the switch adapted to communicate with the switch. The buffer unit is adapted to buffer the data and/or packets until a predefined number of wavelengths leading to a buffered packets destination is/are vacant.

Further the present invention discloses a method for organizing dataflows in an asynchronous communication network including at least one switch and/or Metro access node where said switch and/or Metro access node is associated with at least one buffer and at least a dataflow that can be divided into data packets said dataflow is communicating with the switch and/or Metro access node and the data packets are buffered in the buffer. The buffer unit are buffering the data and/or packets until a predefined number of wavelengths leading to a buffered packets destination is/are vacant.

Brief Description of the Drawings

In order to make the invention more readily understandable, the discussion that follows will refer to the accompanying drawing.

Figure 1 illustrates a generic model of a simulated switch,

Figure 2 is, for a switch, illustrating PLR (Y-axis) as a function of time packets stay in the buffer (X-axis). Clocking in and out of the buffer is not counted into the delay. Some of the plotted points, indicates both high PLR and delay. These points are a result of unfavourable combinations of Wv_1-Wv_3,

Figure 3 illustrates PLR (Y-axis) as a function of number of buffer interfaces (X-axis), Sd = Standard Deviation,

Figure 4 is illustrating Delay in units of duration of mean packet length (units, Y-axis), as a function of number of buffer interfaces (X-axis), Sd = Standard Deviation.

Detailed Description of the Invention

In the following is given a detailed description of a queue arrangement system of general usage; however in the

following description it is described with references to telecommunication and data communication systems. The disclosed arrangement and method is characterised in that it is very flexible in that it employs a feedback buffer, wherein the feedback buffer can be of any kind provided it can be accessed at an arbitrary point of time. Further to accomplish the advantages of the present invention the queue system will be of a type where there is a number of queues and the queues are prioritizing queues where certain parameters is used to decide the amount of queues and/or the dimensioning of the queues. Within telecommunication and/or data communication a natural choice for the criterion parameter is the length of the packets of data and/or the length of the data in other formats such as databursts, data streams or any other dataflow that can be split into packets.

Following the idea of queue organisation of data within a data communication or telecommunication network one can further describe the prioritation of the queues with regard to the length of data packets where a first range of length of packets is associated with a first queue, a second range of packet length is associated to a second queue, a third range of packet length is associated with a third queue and a n'th length of packets is associated with queue number n.

A very effective use of the invention is within networks with optical packet switching employing WDM and where priority according to a set of rules are given regarding data entering a switch from the network, hence an input priority (IP) scheme will be described in details in the following with references to the accompanying drawings.

The described principle is intended to support applications with a high demand to packet sequence and PLR, still saving component costs. This is achieved using electronic buffering in combination with optical switching for optical packet switches. In Metro packet rings, electronic queuing

systems are applied, using the buffering solution according to the present invention.

A first embodiment of the present invention

An approximation to IP can however be done, strongly reducing the contention problem caused by scheduling packets from the buffer. This is achieved when using the proposed AIP3Q algorithm according to the present invention. PLR and reordering of packets is a trade off. To reduce the probability for contention, we can let the packets stay in the buffer until a given number of wavelengths leading to the buffered packets destination are vacant. The higher the number of wavelengths, the lower the probability for congestion. The drawback is that due to the increased delay of the buffered packets, the degree of packet reordering will also increase. A balance between delay and gained PLR must therefore be made.

The short packets will occupy the output-resources for a short time. The probability for the next packet arriving at the switch will be blocked when arriving at a random time after a short packet will therefore be lower than if the previous packet was a long packet. Since buffered short packets introduce a lower probability for contention than the long packets that are buffered, hence according to the present invention the buffered packets is divided into three queues according to the length of the packets. The number of wavelengths that needs to be vacant before scheduling packets from the queue with short packets can be lower than for the queue with the medium length packets, which again has lower demands to number of vacant wavelengths than the queue with the longest packets. The number of minimum vacant wavelengths before a queue is serviced can be defined as: Wv_1, Wv_2 and Wv_3 for the three queues respectively. An empirical Internet packet length distribution like in [1] is assumed, and the packet length ranges for the three queues is defined, Q_1, 40-44 bytes

(B), Q_2 45-576 B and Q_3 577-1500 B. To set Wv_i, a simulation is performed for 32 wavelengths, 8 input fibres, assuming independent input sources and Poisson packet arrival, load 0.8, and set the number of buffer ports to 16. Wv_1-Wv_3 is then varied while always keeping $Wv_1 < Wv_2 < Wv_3$, finding the PLR as a function of mean delay of the buffered packets and the values of Wv_1-Wv_3. A generic model of the simulated switch and the simulation results are shown in fig. 1.

At a buffer delay of two mean packet lengths (three when counting delay caused by clocking), the PLR are close to an asymptote. Hence this value of the delay is chosen and it is found by traversing the simulation data, that Wv1 = 5, Wv2 = 7 and Wv3 = 10.

To emulate the ideal case of total IP, i.e. no extra contention is caused by packets scheduled from the buffer, buffered packets can simply be dropped and not counted in the PLR statistics. This will give the minimum limit, titled IPlim in figure 2, for the achievable PLR in the described system. As a measure of packet reordering, the mean delay of the buffered packets and its Standard Deviation (Sd) is used. In figure 2, the PLR and delay parameters performance for different buffering schemes, also FDL buffering schemes not employing AIP3Q, is shown. The two FDL buffering schemes: INCremental FDL (INC), and FIXed FDL (FIX), both employs buffer priority, since buffered packets are scheduled (or dropped) as soon as they appear at the output of the FDL's. Also the performance of an electronic buffer with Buffer Priority (BP), where packets in the buffer are scheduled as soon as an output becomes available, is found.

In [5], it is suggested that demanding applications will require a PLR of 10^{-6} or better. In the following discussions parameters for a PLR of 10^{-6} which is regarded as sufficient even for demanding real-time applications, is therefore compared. When using a reasonable number of

buffer interfaces, the FIX scheme does not show a sufficiently low PLR. At the performance limit, IPlim, 27 buffer interfaces is required for achieving a sufficient PLR. The electronic BP scheme needs 46 interfaces, which is 70 % higher than the limit, while using AIP3Q, the performance is very close to the limit. Using the INC scheme, 34 buffer interfaces, 26 % more than the limit, is needed. This demonstrates that FDL buffering can be quite effective for reducing PLR in asynchronous packet switching.

However, looking at the delay performance, the INC scheme both shows the highest mean delay and Sd. If these values are added, a value indicating a delay that it is likely to find on some of the buffered packets will be found. For the INC scheme, this value is 14 units. The mean packet length is 286 B. Comparing packets of equal length, assuming they belong to the same application, this implies that in the worst case, some of the shortest packets of 40 B can be passed by more than (286 B/40 B) * 14 units = 100 short packets. This may limit the maximum bandwidth of an application, not tolerating packet reordering, to 1/100 of the links bandwidth. Studying the performance of AIP3Q, the sum of mean delay and its Sd equals 6 units. Doing the same calculation as above, application bandwidth may be limited to 1/43 of the links bandwidth. However, since the queue for short packets have the highest scheduling priority, the figure found must be considered as a conservative value compared to the value for the INC scheme. Looking at the BP scheme, packets can be scheduled as soon as they are clocked in to the buffer and a vacant output is found. The mean time a packet stays in the buffer is then found to be as low as $6.8*10^{-4}$ units, and therefore not plotted in the figure. Hence, packets of the same length may, because of the clocking of the packet into the buffer, be reordered on the link if the same packet is buffered in several preceding switches. Reordering of the packets of an

application is therefore only likely if it demands a bandwidth close to the link bandwidth.

Future applications may have strict demands to both PLR and reordering of packets. Given an asynchronous optical packet switch with limited number of buffer interfaces, the achievable service quality with respect to the mentioned parameters has been evaluated for two electronic and two FDL based buffering schemes. Results show that fixed length FDL's are inefficient, and a sufficiently low PLR cannot be reached. The incremental length FDL's shows good PLR performance, however the figures for delay indicates that reordering of packets is likely to occur if the applications bandwidth demand is equal to or higher than 1/100 of the link bandwidth. When using electronic buffering, a trade off between PLR and packet reordering can be made. The suggested AIP3Q algorithm shows a PLR performance close to an achievable limit, while the critical applications bandwidth can be more than doubled compared with the use of the incremental FDL scheme. If the critical applications bandwidth is close to the link bandwidth, an electronic buffer priority (BP) scheme can be used, but at the cost of adding 70 % more buffer interfaces. When choosing buffering scheme in optical packet switches, in addition to the implementation costs, also the service quality demands from future applications must therefore be carefully considered.

A second preferred embodiment of the invention

In the first preferred embodiment it was described how packet data, data burst etc. where forwarded from the switch input to the number of buffers according to the AIP3Q algorithm.

However, this approach is but one out of a number of approaches where the principles of AIP3Q algorithm according to the present invention can be utilized.

Imagine a system comprising a media access protocol for accessing an asynchronous (metro) packet ring. Following this approach, one will realize that, the only difference in principle is that the packets or databursts will be routed from external lines and directly to the number of buffers. More specifically the input to the buffers will be low bit rate lines - aggregating inputs - which will be aggregated in different queues to the switch' outputs. Hence aggregation of packets in a number of different queues, where each queue, one or more, has packets with a defined packet length associated thereto, for access to a number of outputs, is described according to the present invention. The output can preferably be WDM outputs, and the buffers, or aggregating queues can be of an electronic or optical type. As will be evident from the foregoing the AIP3Q algorithm is a most versatile algorithm for traffic handling within a switch.

Note that while in the foregoing, there has been provided a detailed description of particular embodiments of the present invention, it is to be understood that equivalents are to be included within the scope of the invention as claimed.

References

[1] S. Bjørnstad et al., "A scalable optical packet switch for variable length packets employing shared electronic buffering." ECOC 2002, vol. 3, P.4.7, 2002.

[2] F. Callegati, G. Corazza, C. Raffaelli: "Exploitation of DWDM for optical packet switching with quality of service guarantees", *IEEE J. Sel. Areas Comm.* vol. 20, no. 1, pp. 190-201, January 2002.

[3] M. J. O' Mahony, D. Simeonidou, D. K. Hunter, A. Tzanakaki: "The application of optical packet switching in

future communication networks", *IEEE Comm. Mag.* vol. 3, no. 3, pp. 128-135, March 2001.

[4] C. Develder, M. Pickavet, P. Deemester: "Choosing an appropriate buffer strategy for an optical packet switch with a feed-back FDL buffer" In: *Proceedings of, ECOC 2002,* vol. 3, pp. 8.5.4 (2 pages).

[5] R.J. Gibbens...[et al.]: "*An approach to service level agreements for IP networks with differentiated services*", Statistical Laboratory, University of Cambridge and Internet and Data Networks, British Telecommunications plc., (Article submitted to Royal Society).